1 Filozoficzne problemy sztucznej inteligencji

Sztuczna inteligencja jako dziedzina nauki stawia przed sobą bardzo ambitne cele, czy są one jednak realne? Czy sztuczny system może być naprawdę inteligentny, czy z elektronicznych modeli wyłoni się jakiś umysł? W szczególności czy systemy symboliczne, oparte na wiedzy, są dobrym kandydatem dla stworzenia takich urządzeń? Jaki jest właściwie związek symboli z tym, co one reprezentują? Jak to się dzieje, że symbole reprezentują coś poza samym sobą? Jest to problem filozoficzny wywodzący się jeszcze od Brentano (1874) i dyskutowany do tej pory przez filozofów umysłu. W szczególności problem nabierania znaczenia przez symbole (symbol grounding problem) jest nadal bardzo żywo dyskutowany (Harnad 1990, 1993). W jaki sposób martwe symbole w systemach formalnych nabierają znaczenia? Nie wystarczy do tego system reguł, które pozwalają z jednego zestawu symboli wyprodukować drugi. Jak się wydaje sam system kontrolny, pozbawiony zmysłów, ciała i środowiska, w którym się rozwija nie będzie zdolny do rozwinięcia prawdziwej intencjonalności. Czy będzie natomiast zdolny do wykazywania prawdziwej inteligencji? Być może nie w akceptowanym przez człowieka sensie. Nad niektórymi z tych zagadnień ludzie zastanawiali się już w starożytności, nad innymi jeszcze przed powstaniem sztucznej inteligencji jako nauki. Trudno jest zrozumieć, skąd wzięły się poglądy współczesne bez pobieżnego choćby podsumowania historii rozwoju głównych idei filozofii umysłu.

1.1 Filozofia umysłu a sztuczna inteligencja

Epistemologia, czyli nauka o poznaniu, to jeden z głównych działów filozofii. Drugim bardzo ważnym działem jest **metafizyka**. Arystoteles nazwał tak wszystko to, co zostało do omówienia po zakończeniu spraw dotyczących natury, czyli fizyki. Filozofia umysłu należy zarówno do epistemologii jak i metafizyki. Systemy oparte na wiedzy to podstawa klasycznego podejścia do sztucznej inteligencji. Czym jednak jest wiedza, jaka jest jej natura? Refleksje filozoficzne nad tym problemem już na samym początku doprowadziły do sprzeczności.

Największym zwolennikiem teorii powszechnych zmian był Heraklit, znany ze słynnego powiedzenia „panta rei”, wszystko płynie. Skrajni zwolennicy tej filozofii, np. Kratylos, uznawali z powodów zmienności za niemożliwą wszelką dyskusję. Znaczenie słów nie pozostaje stałe. Zanim wypowiedź dobiegnie końca zmieni się zarówno słuchacz jak i mówca, komu więc mam odpowiedzieć skoro osoba pytająca już się zmieniła? Zwolennicy stałości, wśród których najbardziej znanym był Parmenides, również doszli do skrajnych poglądów. Jeśli istnieje trwała, niezmienna podstawa bytu, to nie może mieć własności dopuszczających zmianę, ruch, dzielenie się, zmianę postaci, może mieć jedynie własność istnienia, nie może więc być częścią dostępnej naszym zmysłom rzeczywistości. „Byt jest, niebytu nie ma”, powiedział Parmenides twierdząc tym samym, że świat zmiennych zjawisk nie istnieje. Zenon z Elei jawnie dowodził, że wszelka zmiana jest niemożliwa. Jego najsłynniejszy paradoks dotyczył żółwia i Achillesa: jeśli żółw znajduje się 10 metrów przed Achillesem to chociaż porusza się dziesięć razy wolniej i tak Achilles nie zdoła go dogonić. Przebycie 10 metrów przez Achillesa oznacza pokonanie w tym samym czasie jednego metra przez żółwia, przebycie kolejnego metra przez Achillesa to 10 cm żółwia, 10 cm Achillesa to 1 cm żółwia i tak w nieskończoność. By przebyć całą drogę trzeba przebyć połowę, by przebyć połowę trzeba połowę tej połowy ... Jeśli strzała mogłaby się poruszać to musi się poruszać tam gdzie jest albo tam gdzie jej nie ma. Tam gdzie jest nie może się jednak poruszać, bo jest to jeden punkt. Tam gdzie jej nie ma jej nie ma, nie może się więc poruszać. Filozofia już na samym początku swojego rozwoju uwikłała się w sprzeczności.

Sprzeczności te rozwiązano na kilka sposobów. **Demokryt** założył, że istnieją trwałe i niezmienne atomy, mające jako niezniszczalne obiekty rzeczywistą naturę. Zderzenia tych obiektów ze sobą w ich wiecznym ruchu w pustce wywołują wrażenie nieustannych zmian. Epikur wprowadził szczególny rodzaj doskonałych atomów sferycznych, zwanych **atomami duszy**. Metafizyka epikurejska była bardzo konsekwentna i tłumaczyła wszystkie zdarzenia przez ruch atomów odbywających się bez żadnego celu.

Pierwszą teorię wiedzy w europejskiej filozofii skonstruował **Platon**. Według niego prawdziwa wiedza dotyczy obiektów idealnych, niezmiennych kategorii ogólnych a nie samych ulotnych form, z którymi mamy na co dzień do czynienia. Język nie odnosi się tylko do konkretów, lecz przede wszystkim do kategorii ogólnych. Rzeczy konkretne są jedynie przykładami pojęć ogólnych. W ten sposób Platon zapoczątkował trwające do dziś w psychologii poznawczej dyskusje i prace nad sposobem rozumienia i tworzenia kategorii. Jego własne

rozwiązanie było dość proste: pojęcia ogólne, nazywane uniwersaliami, powszechnikami lub ideami Platońskimi, istnieją naprawdę. Nie można ich pojąć przez doświadczenie zmysłowe a jedynie przez kontemplację. Wiedzy nie można zdobyć przez naukę, albowiem wiedzę osiąga się przez rozpoznanie. Widzę coś i mogę to poznać jedynie dzięki temu, że już to znam. W takim razie cała wiedza, którą mozolnie zdobywamy ucząc się, musi już w naszych umysłach tkwić, tylko o niej zapomnieliśmy. Nauka jest jedynie przypominaniem i rozpoznawaniem. Filozof drogą rozmyślań zbliżyć się może do poznania świata idei. Najważniejszymi naukami są więc arytmetyka i geometria, pozwalające dostrzec blask idealnych form. Według Sokratesa „arytmetyka posiada wielki efekt wyzwalający i uwznioślający, skłaniający duszę do zastanowienia się nad abstrakcyjnymi liczbami i buntowania się przeciwko wprowadzaniu postrzegalnych czy też dotykalnych obiektów argumentacji". Kolejnym etapem studiów na drodze do świata idei jest astronomia oraz studiowanie harmonii dźwięków, znacznie rozwinięte przez szkołę Pitagorejską. Po takim przygotowaniu filozof gotowy jest do ostatecznego kroku na drodze do wyzwolenia z królestwa cieni, a jest nim studiowanie **dialektyki**. Nauka dialektyki jest dość mglisto opisana jako „odkrywanie absolutu jedynie dzięki światłu rozumu, bez posługiwania się zmysłami, nie ustając póki dzięki czystej inteligencji nie dotrze się do absolutnego dobra". Świat zjawisk jest niedoskonały i można o nim jedynie snuć domysły, prawdziwa pewna wiedza możliwa jest jedynie w odniesieniu do świata rzeczywistego, jakim jest świat idei. Fizyka bada jedynie świat cieni a matematyka świat rzeczywisty.

Kartezjusz (1596-1650**)** był pierwszym wybitnym dualistą, oddzielając subiektywne od obiektywnego, umysł od ciała, poznającego od poznawanego, torując drogę mechanicystom. Newton udoskonalił ten schemat: naszym zadaniem jest odkrycie planów budowy Boskiego Zegarka, jakim jest natura. Badając „jasne i wyraźne" idee dotyczące swojego umysłu Kartezjusz doszedł do wniosku, że musi się on składać z innej substancji niż obiekty fizyczne. Myśli nie mają takich cech jak wielkość czy kształt, charakteryzująca ciała fizyczne. Ciała fizyczne są rozciągłe i mają liczne własności, których nie można przypisać myślom. Ponieważ działanie wymaga kontaktu, w jaki sposób kontaktuje się umysł ze światem fizycznym? Oddziaływanie świata na umysł jest faktem znanym każdemu z doświadczenia. W samym środku mózgu znajduje się gruczoł dokrewny zwany szyszynką. Tu właśnie upatrywał Kartezjusz łączności pomiędzy światem fizycznym i psychicznym, chociaż zdawał sobie sprawę, iż nie jest to dobre rozwiązanie. Trudno jest bowiem ustalić, czy szyszynka ma naturę fizyczną czy mentalną. W jednym z listów Kartezjusz przyznał, że jest to jedna z tajemnic, której prawdopodobnie nie da się wyjaśnić.

Kartezjusz sformułował problem psychofizyczny, czyli problem relacji ciała i umysłu, ale nie utrzymywał, że potrafi go rozwiązać. Thomas Hobbes, jego zawzięty krytyk, tak pisał do swojego oponenta: „Umysł to nic innego jak ruchy pewnych części ciała organicznego". Według Hobbes'a „rozumowanie to nic więcej niż kombinowanie" a więc myśl powstaje w wyniku obliczeń, zastosowania formalnych operacji do reprezentacji należących do świata umysłu. Kartezjusza i Hobbsa uznaje się za prekursorów nauk o poznaniu w ich kognitywistycznym wydaniu. W 1976 roku pojawiła się praca Allana Newella i Herberta Simona będąca manifestem kognitywizmu. Umysł to maszyneria służąca przetwarzaniu informacji, a przetwarzanie informacji to manipulacja symbolami, a więc coś, co komputery lubią najbardziej. Wynika stąd, że komputery nie tylko symulują inteligentne zachowanie ale też naprawdę myślą i mają stany poznawcze. Pogląd taki stał się bardzo popularny wśród specjalistów od sztucznej inteligencji i zwykle określany jest jako „silne AI" (strong AI), czyli silna wersja przekonania głosząca, że odpowiednio zaprogramowany komputer jest równoważny umysłowi.

James Mill pisał: „Chcę zniszczyć złudzenia psychicznej aktywności, zredukować wszystko do stałych i w jakimś sensie mechanicznych związków elementów, które powinny być możliwie najprostsze." Był to mechanicyzm w czystej postaci - krzyki i jęki to nie wyraz cierpienia lecz podobne zgrzyty, jak wydają źle naoliwione maszyny. Na tym gruncie szybko przyjęła się teoria Darwina. Najsilniejsi zwyciężają - dobre uzasadnienie dla rodzącego się kapitalizmu bez skrupułów. Była to wizja jeszcze bardziej degradująca rolę umysłu. W 1902 roku Pawłow odkrył warunkowe refleksy i chociaż sam ostrzegał przed generalizacją amerykański psycholog John Watson, a potem B. Skinner, stworzył **behawioryzm**, przez wiele lat dominującą w amerykańskim społeczeństwie psychologię bez umysłu.

Poglądy Kartezjusza i innych filozofów racjonalistycznych spotkały się z ostrą krytyką ze strony **empirystów**, twierdzących, że to właśnie dane zmysłowe są źródłem wszelkiego poznania. Naukowcy tego okresu, Izaak Newton, Robert Boyle i Robert Hooke, dokonali olbrzymich postępów dzięki zastosowaniu empirycznej, opartej na wynikach eksperymentów, metody badania świata. Nic dziwnego, że i filozofowie, poczynając od Francisa Bacona, zaczęli uznawać badania empiryczne za ważne źródło wiedzy o świecie i o nas samych. Doktor medycyny John Locke (1632-1704), autor dzieła *Rozważania dotyczące rozumu ludzkiego*, dowodził, że cała wiedza powstaje dzięki zmysłom i nie mamy żadnych idei wrodzonych. Ani dzieci, ani upośledzeni umysłowo, nie zdają sobie sprawy z idei, uznawanych przez racjonalistów za wrodzone. Człowiek rodzi się jak biała kartka

papieru, „tabula rasa”, i dopiero doświadczenie zmysłowe i refleksję nad nim tworzy idee, na których opieramy swoje myślenie o świecie. Proste idee wiążą się bezpośrednio z doświadczeniem zmysłowym. Umysł gromadzi, powtarza i łączy te podstawowe idee tworząc idee złożone. Dyskusje nad tym, czy istotnie wszystkie koncepcje umysłowe pochodzą z doświadczenia zmysłowego, czy też mogą być wytworem samego umysłu, nie zakończyły się do dzisiaj. Ich echem jest dyskusja zmierzająca do rozstrzygnięcia, czy czynniki genetyczne czy też środowiskowe są najważniejsze dla kształtowania się umysłu, a w szczególności charakteru człowieka.

Jednym z najbardziej znanych filozofów umysłu był szkocki sceptyk David Hume (1711-1776). Jego dwa podstawowe dzieła, napisane w młodym wieku, to *Traktat o naturze ludzkiej* i *Badania dotyczące rozumu ludzkiego.* Ambicją Hume'a było zastosowanie metody naukowej Newtona do zbadania natury umysłu i odkrycie, w jaki sposób powstają nasze poglądy, można go więc uznać za prekursora psychologii. W umyśle dostrzegamy wrażenia i idee, różniące się stopniem, w jakim przyciągają naszą uwagę. Idee i wrażenia proste „nie dopuszczają oddzielenia”. Idee złożone nie zawsze pojawiają się jako wrażenia ale wszystkie idee proste są związane z prostymi wrażeniami. Hume wyciągnął stąd wniosek, że nie ma idei wrodzonych, wszystkie idee proste powstają z prostych wrażeń a złożone z idei prostych. Pamięć przechowuje idee w ustalonym porządku, wyobraźnia porządkuje je według życzeń. W procesie myślenia można dostrzec swojego rodzaju przyciąganie, podobne do Newtonowskiej grawitacji. Zdobywamy wiedzę w sposób intuicyjny i pewny porównując ze sobą idee w umyśle. Umysł działa opierając się na skojarzeniach, czy to czysto przyczynowych czy związanych z następstwem czasowym lub miejscem. Chociaż wszystko, co widzimy, to tylko bezpośrednie następstwa zdarzeń lub ich sąsiedztwo nie wystarcza to do określenia przyczyn danego zjawiska. Błyskawica poprzedza grzmot, nie oznacza to jednak, że jest jego przyczyną. Jeśli doznajemy jednocześnie kilku wrażeń to mamy skłonność do kojarzenia ich razem i uznawania je za trwale połączone. Robimy więc ukryte założenie o jednolitości, czy też powtarzalności, przyrody. Wszystkie prawa przyrody zależą od tej zasady jednolitości lecz jej samej nie potrafimy w żaden sposób udowodnić. Chociaż jest to bardzo nieprawdopodobne prawo grawitacji może jutro przestać działać. Nie ma powiązań koniecznych pomiędzy rzeczami, wiara w przyczyny wynika tylko z przyzwyczajenia umysłu. „Efektywność przyczyn leży w determinacji umysłu”, stwierdził Hume. Związek pomiędzy zdarzeniami ma więc naturę psychologiczną. Takie rozważania doprowadziły Hume'a do totalnego sceptycyzmu. Prawa nauki nie mówią nic o przyrodzie a jedynie o nawykach naszego umysłu.

William James (1842-1910) był nie tylko wybitnym filozofem pragmatykiem lecz również ojcem nowoczesnej psychologii. Zakwestionował on potrzebę opierania naszej wiedzy na absolutnie pewnych podstawach. Zgodnie z Kantem nasza wiedza wynikająca z doznań zmysłowych jest interpretowana w oparciu o wrodzone nam kategorie. Nie mamy jednak gwarancji, że wrodzone kategorie odpowiadają jakiejś prawdzie. Teorie są jedynie narzędziami do rozwiązywania problemów i nie należy ich oceniać jako prawd absolutnych a jedynie na podstawie skuteczności w rozwiązywaniu problemów. Ponieważ przyjęcie lub odrzucenie wielu klasycznych poglądów i teorii filozoficznych nie ma wpływu na nasze działanie teorie te nie mają żadnej wartości, nie pomagają w rozwiązaniu żadnego konkretnego problemu, który przed nami stoi. Teoria jest prawdziwa, jeśli jest skuteczna. Prawda nie jest jakąś własnością niezależną od ludzkiego doświadczenia. To, co potwierdza się w doświadczeniu uznajemy za prawdziwe. Z idei pragmatyzmu najbardziej konsekwentne wnioski wyciągnął John Dewey tworząc **instrumentalizm**. Myślenie jest instrumentem pozwalającym rozwiązywać problemy. Wynikały z tego praktyczne skutki, szczególnie dla edukacji. Zamiast zmuszać dzieci do zapamiętywania niepotrzebnych faktów należy je uczyć rozwiązywania problemów. Idee Deweya wywarły i do tej pory wywierają duży wpływ na szkolnictwo amerykańskie.

Współczesną logikę w znacznym stopniu stworzył Gottlob Frege (1848-1825), ale dopiero pojawienie się wielkiego dzieła *Principia Mathematica* Bertranda Russela i Alfreda Whiteheada (w trzech tomach, wydanych w latach 1910-1913) wywarło duży wpływ na filozofię. Logika symboliczna zawierała klasyczną logikę Arystotelesa jako niewielką część, pokazywała również w jaki sposób wiele działów matematyki daje się zredukować do logiki i teorii zbiorów. Nowa logika odnosiła się do zdań i relacji pomiędzy zdaniami. Russell był przekonany, że logika zdań wystarczy do zrozumienia języka potocznego, do określenia znaczenia zdań a tym samym doprowadzi do odkrycia praw myślenia. *Principia Mathematica* daje nam schemat języka doskonałego, odzwierciedlającego strukturę rzeczywistego świata. Nareszcie marzenia Raymonda Lulla, Leibniza i wielu innych uczonych na przestrzeni wieków miały się spełnić: stworzono rachunek logiczny pozwalający rozstrzygnąć kwestie filozoficzne.

W najpełniejszy sposób idee atomizmu logicznego przedstawił Ludwig Wittgenstein w *Traktacie Logiczno-Filozoficznym*, opublikowanym w 1921 roku. Według niego język idealny odzwierciedla świat, jest pewnego rodzaju mapą odbijającą strukturę rzeczywistości, a więc pokazuje nam strukturę faktów dotyczących obiektów i ich własności. Nauki empiryczne opisują nowe fakty a filozofia zajmuje się strukturą świata. Analiza

filozoficzne polega na analizie sensu zdań przez przedstawienie ich w postaci formuł logicznych, stąd ten rodzaj filozofii można nazwać **filozofią analityczną**. Atomizm logiczny nie jest jedyną formą filozofii analitycznej a jego największa popularność przypadła na okres międzywojenny.

W latach dwudziestych grupa żyjących w Wiedniu uczonych, którzy przeszli do historii pod nazwą Koła Wiedeńskiego, wyciągnęła jeszcze dalej idące wnioski z logicznych odkryć Russella i Whitteheada. Już Wittgenstein w swoim *Traktacie* zauważył, że filozofia nie jest teorią ale pewną działalnością zmierzającą nie tyle do odkrycia nowej wiedzy lecz raczej lepszego zrozumienia wypowiedzi. Zrozumienie umysłu wymaga przede wszystkim zrozumienia języka. Umysł należy uznać za maszynę do wnioskowań logicznych operującą na symbolach reprezentujących idee, szczególnie symbolach języka. Jak wielki był wpływ tych idei świadczyć może fakt, że jeszcze w latach 70-tych specjaliści od sztucznej inteligencji próbowali zastosować logiczną reprezentację wiedzy nawet w takich zagadnieniach jak analiza obrazu! Rezultaty nie były jednak zachęcające i logiczna reprezentacja wiedzy powoli traciła na znaczeniu. Logiczne podejście do teorii języka odrzuciło wielu filozofów, w tym sam Wittgeinstein, którego *Dociekania Filozoficzne*, opublikowane w 1952 roku już po śmierci autora, wywarło wielki wpływ na filozofię współczesną. Wittgenstein opowiedział się za filozofią języka potocznego uznając, że logika nie wyczerpuje możliwości języka. Nie ma jednak wątpliwości, iż logika w dalszym ciągu dla wielu ludzi jest najważniejszym drogowskazem. Mogłem się o tym przekonać osobiście korespondując w 1994 roku z pewnym amerykańskim logikiem który twierdził, iż naprawdę wyjaśnił strukturę umysłu i świadomość. Oczywiście nie wyjaśnił żadnego konkretnie obserwowalnego empirycznie zjawiska, tylko „świadomość w ogóle”.

Problem ciała i umysłu lub zagadnienie psychofizyczne to jeden z wielkich tematów filozofii umysłu. W jaki sposób odczuwane przez nas wrażenia wiążą się z rzeczywistością? Najbardziej skrajnym rozwiązaniem filozoficznym jest **solipsyzm**, czyli pogląd, że tylko ja istnieję, a moje wrażenia i wszyscy inni są tylko produktami mojej wyobraźni. **Dualizm** wywodzący się od Kartezjusza zakłada całkowitą odrębność świata umysłu i świata ciała (zjawisk fizycznych). **Idealizm** jest skrajnym przeciwieństwem materializmu i utrzymuje, że wszystko ma charakter psychiczny. Teoria **monad** Leibniza głosi, że każda jednostka obdarzona własnościami fizycznymi i psychicznymi jest odrębnym, niezależnym bytem tworząc **monadę**. Wszystko co się jej zdarza wynika z właściwości jej wewnętrznej natury i chociaż nie jest zależne od innych dzięki istnieniu synchronizacji, nazywanej przez Leibniza „odwieczną harmonią”, jej wrażenia odpowiadają zdarzeniom w innych monadach lub fizycznych przedmiotach. Podobne rozwiązanie problemu ciała i umysłu znalazł Spinoza. Chociaż między zdarzeniami psychicznymi i fizycznymi nie ma bezpośredniego związku istnieje **paralelizm**. Umysł i ciało są atrybutami tego samego bytu. Świat mentalny i fizyczny to dwie strony natury lub Boga, jak kto woli. Porządek logiczny umysłu odpowiada fizycznemu porządkowi w przyrodzie.

Ponieważ dualizm ani idealizm nie zrobiły w ostatnich stuleciach żadnych postępów wielu badaczy skłonna jest uznać **materialistyczną metafizykę**. W jej ramach mieści się wiele nurtów. W psychologii **behawioryzm** uznał, że takie zagadnienia jak umysł czy świadomość nie istnieją, warto zajmować się jedynie zdarzeniami fizycznymi, odruchami i impulsami nerwowymi. **Epifenomenalizm** głosi, że myśli i wrażenia to specjalne stany, z niejasnych przyczyn będące ubocznymi skutkami fizycznych stanów mózgu, ruchów materii. **Funkcjonalizm** głosi, że realizacja procesów przetwarzania informacji w konkretnej, biologicznej postaci, jest rzeczą mało istotną. Umysł jest rezultatem procesów obliczeniowych i jeśli odpowiednie procesy będą zachodzić w umysłach istot zrobionych z krzemu to będą one miały umysł, niezależnie od konstrukcji ich mózgu. Proste procesy, nie posiadające żadnej inteligencji, współdziałając razem potrafią rozwiązywać coraz bardziej złożone zadania. Inteligencja wyłania się więc drobnymi kroczkami z hierarchicznej struktury programów w sztucznej inteligencji. Problemem nie jest kwestia symulacji inteligencji, gdyż nie ma wątpliwości, że w mniejszym lub większym stopniu jest to możliwe. Problemem jest kwestia intencjonalności, zdolności procesów algorytmicznych do odczuwania i rozumienia sensu symboli, z którymi mają do czynienia.

1.2 Kognitywizm

Idee kognitywistyczne powstały wśród specjalistów zajmujących się sztuczną inteligencją i psychologią poznawczą. Komputery pokazały, że możliwe jest celowe i inteligentne zachowanie się złożonego systemu pomimo tego, że żaden z elementów nie jest sam w sobie inteligentny. Dzięki temu odpadł argument Gilberta Ryle'a dowodzący, że do osiągniecia inteligencji potrzebny jest „duch w maszynie”, gdyż każdy inteligentny proces musi się odwoływać do równie inteligentnych podprocesów. Inteligentne zachowanie może być wynikiem kooperacji wielu prostszych podprocesów, z których każdy rozbić można na jeszcze prostsze podprocesy itp. Komputery przetwarzają informację w formie symbolicznej. Symbole mogą być zapisane w różnej formie, np.

śladów atramentu na kartce papieru, stanu magnetyzacji dysku czy impulsów elektrycznych, lecz reprezentują one coś innego niż swoja fizykalną formę. Komputer manipuluje symbolami w oparciu o pewne reguły, które określają znaczenie symboli w oparciu o ich wzajemne związki. Systemy przetwarzające informację kodują ją w postaci symboli, funkcjonalnie zorganizowanych i stojących względem siebie w określonych przez dane reguły relacjach. Stany umysłowe, takie jak myśli, przekonania, procesy percepcji można więc uznać za obliczeniowe (realizowane przez różne stany fizyczne) maszynerii przetwarzającej informację opisanej w sposób funkcjonalny. Ten sam stan umysłowy może być zrealizowany w różny sposób w różnych systemach fizycznych; te same stany fizyczne w odmiennie zorganizowanych systemach przetwarzania informacji mogą odpowiadać różnym stanom umysłowym.

Funkcjonalizm kładzie więc nacisk na funkcję uznając sposób fizycznej realizacji samego procesu reprezentacji symboli za sprawę drugorzędną. Istoty organiczne mają biologiczne mózgi, istoty krzemowe mogą mieć mózgi elektroniczne a jeszcze inne istoty czysto optyczne. Nawet niewielki komputer może symulować działanie potężnego superkomputera przyjmując te same stany procesora, chociaż znacznie wolniej. Jest to pogląd dużo bardziej wyrafinowany niż behawioryzm - przetwarzanie informacji nie da się sprowadzić do prostych asocjacji - a jednocześnie nie zakładający dualizmu substancji ciała i umysłu. W tym ujęciu umysł jest funkcją procesów fizycznych zależną od organizacji przetwarzania informacji, a więc od algorytmu realizowalnego przez mózgową maszynerię. Skoro tak to umysł powinien dać się zrealizować również przy pomocy komputera jeśli tylko znajdziemy na to odpowiedni algorytm. Odpowiednio zaprogramowany komputer powinien być równoważny umysłowi i powinien mieć stany poznawcze. To, że na razie nie ma takich komputerów nie jest żadnym argumentem. Przyzwyczailiśmy się do głupich maszyn, ale to nie znaczy jeszcze, że maszyny takie muszą być.

Przetwarzanie informacji przez człowieka na pewno w pewnym zakresie można uznać za proces dyskretny, w odróżnieniu od przetwarzania analogowego. Symbole językowe są elementami dyskretnymi, nie zmieniają się płynnie przechodząc jeden w drugi. Jednocześnie pewne procesy percepcji wydają się być procesami analogowymi, np. wyobraźnia przestrzenna. Umysł jest wiec rezultatem procesów hybrydowych, analogowo-dyskretnych. Funkcjonalizm skupia się na procesach złożonych, na wyższych czynnościach poznawczych wymagających używania języka i rozumowania. Intencjonalność systemu przetwarzania informacji, czyli uznanie, że symbole, na których operuje, odnoszą się do czegoś, są istotnie o czymś, wymaga pewnego izomorfizmu relacji pomiędzy obiektami w świecie rzeczywistym i w świecie umysłu. Ulubioną reprezentacją struktury relacji pomiędzy obiektami w sztucznej inteligencji stały się sieci semantyczne i hierarchiczne drzewa pojęć. Nie wystarcza to jednak do osiągnięcia intencjonalności, wydaje się że konieczna jest tu jeszcze odpowiednia zależność przyczynowa od stanów świata. Zdarzenia wewnętrzne muszą wywoływać odpowiednie reprezentacje. Nawet jeśli nie wystarczy to do osiągnięcia intencjonalności sztucznego systemu jest to warunek konieczny. Jaki jest właściwie związek symboli z tym, co one reprezentują? Jak to się dzieje, że symbole reprezentują coś poza samym sobą? W jaki sposób martwe symbole w systemach formalnych nabierają znaczenia?

Nauki o poznaniu do początków lat 90-tych oparte były przede wszystkim na uznaniu natury umysłu za system przetwarzania informacji i próbie stworzenia takich modeli umysłu. W dalszym ciągu niektórzy przedstawiciele nauk o poznaniu są przekonani, że właśnie w ten sposób uda się zrobić sztuczny umysł. Zagadnienie jest bez wątpienia bardzo skomplikowane i wymaga ogromnych baz wiedzy i rozwinięcia lepszych metod reprezentacji wiedzy oraz sposobów korzystania z takiej wiedzy. Najpoważniejszym obecnie projektem jest prowadzony od 1984 roku przez Douglasa Lenata projekt CYC, zmierzający do stworzenia programu obdarzonego zdrowym rozsądkiem. Wśród innych dużych projektów zmierzających do stworzenia sztucznego umysłu i czerpiących swoje nadzieje na powodzenia tego przedsięwzięcia z funkcjonalizmu wymienić należy projekt SOAR Allana Newella i projekt OSCAR Johna Pollacka. Niezależnie od wyników takich prób - trzeba przyznać, że jak dotychczas nie są one zbyt imponujące - powstaje poważny problem epistemologiczny. Czy systemy komputerowe symulują jedynie inteligentne działanie czy tez naprawdę mogą być równoważne umysłowi i mają stany poznawcze? Kognitywizm czerpie swoje przekonanie o tym, że jest to właściwa droga do stworzenia modeli sztucznego umysłu z funkcjonalizmu.

Jednym z najwybitniejszych przedstawicieli kognitywizmu jest Allen Newell. Do spółki z Herbertem Simonem sformułował on, a może raczej tylko bardziej unaocznił, podstawy całej dziedziny. **Umysł jest systemem kontrolnym** określającym zachowanie się systemu w jego skomplikowanych oddziaływaniach ze środowiskiem. Umysł dostarcza różnorakich funkcji określających odpowiedzi organizmu na sytuacje środowiska. Dla każdej sytuacji (lub dla każdego typu sytuacji) odpowiedzi te mogą być różne, zależne nie tylko od aktualnego stanu środowiska, ale i od historii poprzednich oddziaływań. Język systemów kontrolnych pozwala przypisać im

pewne cele. Realizacja tych celów wymaga posiadania wiedzy. Umysł jest systemem kontrolnym posiadającym liczne cele i wykorzystującym szeroką wiedzę. Systemy można opisywać na różnych poziomach, od składu molekularnego substancji z której są fizycznie zrobione do poziomu organizacji ich zachowania. Mówiąc o zachowaniu używamy języka intencji, celów, wiedzy. Na poziomie programu mówimy o instrukcjach, na poziomie sprzętowym mówimy o zachowaniu się sterowanych urządzeń, na poziomie opisu fizycznego o ich własnościach fizycznych. Poziom funkcjonalny opisu organizacji i zachowania jest według kognitywistów poziomem intencjonalnym. Różne systemy intencjonalne redukują się na różnych poziomach do poziomu molekularnego opisywanego przez prawa fizyki (por. Tabela). Każdy z tych poziomów różni się jakościowo od siebie. Poziom intencjonalny jest po prostu jednym z poziomów opisu systemów działających w oparciu o wiedzę.

Newell twierdzi, że używanie takich pojęć jak: „ten program już to wie” nie jest to tylko metafora językowa ale stwierdzeniem coraz bardziej bliskim prawdy. John McCarthy, twórca terminu „sztuczna inteligencja” i najbardziej w tej dziedzinie rozpowszechnionego języka programowania LISP twierdzi, że nawet tak proste urządzenia jak termostaty mają przekonania: za gorąco, za zimno, odpowiednio. To, co kognitywiści nazywają

Poziom:	Systemy przetwarzające wiedzę	Umysły
Substrat:	Wiedza	Świat wewnętrzny
Prawa:	Zasady racjonalnego działania	Prawa psychologii
Poziom:	Systemy oprogramowania	Zachowania wyuczone ?
Substrat:	Struktury danych i programy	Neurodynamika
Prawa:	Interpretacja syntaktyczna instrukcji	Dynamika złożonych układów
Poziom:	Uniwersalny komputer	Mózg
Substrat:	Ciągi bitów	Stany neuronów
Prawa:	Arytmetyka binarna	Reguła Hebba
Poziom:	Architektura sprzętowa	Przetwarzanie różnorodnych sygnałów
Substrat:	Obwody logiczne	Funkcjonalne grupy neuronów
Prawa:	Logika, reguły projektowania	Neurofizjologia, geny
Poziom:	Obwody elektryczne	Neurony
Substrat:	Napiecia/prądy/zjawiska elektryczne	Zjawiska elektryczne
Prawa:	Prawa Ohma, Kirchoffa, Faradaya	Prawa Ohma, Kirchoffa, Faradaya
Poziom:	Obwody scalone	Biochemiczny
Substrat:	Atomy i elektrony półprzewodników	Neurocząsteczki
Prawa:	Fizyka ciała stałego	Fizyka molekularna

wiedzą filozofowie skłonni są raczej określać jako przekonania czy wierzenia, bo wiedza w systemie SOW nie musi być związana z prawdą.

Reprezentacja wymaga pewnego substratu, materiału, struktur, które mogą znaleźć się w dowolnie dużej liczbie konfiguracji, dzięki czemu będą mogły opisać dowolnie złożone sytuacje rzeczywiste. Systemy intencjonalne mogą korzystać z różnych sposobów reprezentacji wiedzy, niewidocznych na poziomie wiedzy. Jedną z ulubionych form reprezentacji wiedzy w filozofii jest oczywiście logika, jednakże nie zawsze jest to reprezentacja wygodna, stad w nauce o sztucznej inteligencji wymyślono bardzo wiele innych sposobów reprezentacji wiedzy. Rzeczą ważną jest zachowanie istotnych relacji pomiędzy elementami reprezentującymi a tym, co jest reprezentowane. Można uznać, że reprezentacja jest pewnym szyfrem zakodowującym informację o istniejącej sytuacji, umożliwiającym transformacje zaszyfrowanej informacji i następnie jej odszyfrowanie, przez co umożliwia przewidywanie zdarzeń w domenie wyjściowej. Istotną cechą dobrych reprezentacji jest możliwość składania transformacji w kombinatoryczny sposób. Wymaga to właściwego substratu dla reprezentacji, zdolnego do tworzenia dowolnie złożonych konfiguracji. Zdolność do reprezentacji danego systemu wymaga składania, kombinacji różnych reakcji systemu. Stabilność substratu, w którym tworzy się reprezentacje, musi być odpowiednio duża, by pozostały ślady pamięci, a jednocześnie odpowiednio mała by dokonać transformacji przy niewielkim nakładzie kosztów energetycznych. Łatwość tworzenia pożądanych transformacji reprezentacji jest jedną z najważniejszych wyróżniających reprezentacje dobre od złych.

System oparty na wiedzy (SOW) oddziaływuje ze środowiskiem, w którym się znajduje, wykonując pewne akcje, które składają się na opis jego zachowania. Wiedza jest substratem przetwarzanym przez system określającym cele jego działania. SOW działa w oparciu o jedno ogólne prawo określające jego zachowanie: podejmować działania by spełnić swoje cele korzystając przy tym w pełni z posiadanej wiedzy. Wewnętrzne przetwarzanie reprezentacji i złożone reakcje wymagają przetwarzania danych, a to robią najlepiej systemy dokonujące obliczeń, czyli systemy komputerowe. Uniwersalne systemy obliczeniowe, znane pod nazwą maszyn Turinga, są teoretycznym modelem współczesnych komputerów. Są one zdolne do realizacji prawie dowolnych transformacji, być może nawet z punktu widzenia modeli układów poznawczych zbyt wielu.

Reprezentacje dają się ująć w postaci symbolicznej. Symbol to pojęcie abstrakcyjne, klasa abstrakcji wszystkich znaków symbolicznych, które reprezentują to samo. Znaki symboliczne wskazują na pewne reprezentacje ale wiedza zawarta jest również w relacjach pomiędzy znakami, w ich wzajemnym położeniu, np. w szyku słów w zdaniu. Systemy symboliczne są jedna z realizacji uniwersalnych systemów komputerowych. Zawierają one pamięć, w której znajdują się struktury złożone ze znaków symbolicznych; symbole będące powtarzającymi się wzorcami jakiegoś substratu, wskazujące na elementy pamięci lub inne struktury; operacje, czyli procesy działające na strukturach symbolicznych i produkujące inne struktury symboliczne oraz procesy prowadzące od struktur do zachowania się systemu, czyli procesy interpretujące struktury symboliczne z punktu widzenia zachowania się systemu. Systemy symboliczne pozwalają urzeczywistnić praktyczne realizacje systemów opartych na wiedzy. Wiedza odnosi się do jakiejś domeny a systemy symboliczne pozwalają na jej dobrą aproksymacje. Symbole reprezentują obiekty i relacje odnoszące się do świata zewnętrznego w stosunku do systemu. Semantyczne znaczenie symboli jest wiec wynikiem oddziaływania systemu ze środowiskiem. Sztuczna inteligencja poszukuje przybliżeń do SOW w oparciu o systemy symboliczne. Architektura systemu to pewna struktura całości realizująca działanie systemu symbolicznego.

Inteligencja nie jest pojęciem dobrze sprecyzowanym, bardzo kontrowersyjne są kwestie mierzenia ilorazu inteligencji usiłujące uśrednić różne rodzaj zdolności i inteligentnych zachowań. Ostatnie badania zmierzają raczej w stronię zdefiniowania siedmiu różnych współczynników określających stopień inteligencji człowieka. W kontekście nauki o sztucznej inteligencji możliwe jest dość precyzyjne zdefiniowanie inteligencji: wszystkie zadania, których nie można rozwiązać przy pomocy algorytmów, wymagając inteligencji. System jest inteligentny o tyle, o ile jest dobrym przybliżeniem SOW, który może rozwiązać postawione przed nim zadanie. Bardziej techniczne użycie słowa inteligencja w kontekście systemów SOW prowadzi do następujących wniosków: jeśli system używa całej dostępnej mu wiedzy i wyciąga z niej wszystkie wnioski to jest doskonale inteligentny. Jeśli systemowi brakuje wiedzy to niezdolność do rozwiązania postawionego przed nim zadania nie jest winą braku jego inteligencji lecz braku wiedzy. System automatycznego tłumaczenia tekstów, który nie zna jakiegoś zwrotu idiomatycznego narobi oczywistych błędów, podobnie jak i człowiek usiłujący zrozumieć dany zwrot (niestety, zbyt często spotykamy się z tego typu problemami słuchając tłumaczeń programów telewizyjnych). Jeśli system posiada wiedzę ale nie potrafi jej użyć to jest to wynik braków jego inteligencji. W tym ujęciu inteligencja sprowadza się do zdolności używania wiedzy do osiągania stojących przed systemem celów. Inteligencja zależy od wiedzy, zależy również od celów: w osiąganiu jakiegoś celu system może wykazywać doskonałą inteligencję a w osiąganiu innych celów zerową. Pojęcie inteligencji daje się zastosować tylko do systemów opartych na wiedzy.

Inteligentne zachowanie wymaga rozważenia pewnej liczby możliwych rozwiązań czy sposobów postępowania, oceny strategii i wyboru najlepszej. Podstawą inteligentnego zachowania są więc procesy poszukiwania rozwiązań. Szukanie jest podstawową metodą informatyki a badania nad sztuczną inteligencją bardzo mocno rozwinęły naszą wiedzę o algorytmach szukania i sposobach optymalizacji problemów. Wiedza o procesach szukania wywarła również duży wpływ na badania psychologiczne związane z wyższymi czynnościami poznawczymi: rozumowaniem, rozwiązywaniem problemów, myśleniem. Podstawowa zasada brzmi następująco: jeśli nie wiadomo, w jaki sposób osiągnąć dany cel utwórz przestrzeń różnych możliwości i przeszukuj ją w celu znalezienia drogi do celu. Jeśli wiemy, jak dany cel osiągnąć, możemy zastosować odpowiednią procedurę obliczeniową, jeśli natomiast nie wiemy, musimy rozważyć duża możliwości, posługując się na każdym kroku dostępną w danym momencie wiedzą. Tworzenie przestrzeni rozwiązań problemów wymaga ustalenia odpowiedniej reprezentacji samego problemu i celów, jakie sobie stawiamy, np. jeśli mamy do czynienia z grami planszowymi o określonych regułach i zgromadziliśmy wiedzę o różnych strategiach przydatnych do osiągnięcia celu to musimy ustalić, która z tych strategii doprowadzi do największej przewagi.

Jednym z nieporozumień dotyczących metod sztucznej inteligencji jest przekonanie niektórych jej krytyków, że komputery zawsze przeszukują wszystkie możliwości a ludzie od razu dokonują świadomych (a przez to w

tajemniczy sposób niepojętych dla komputera) wyborów optymalnych możliwości. Komputer rozważa miliony wariantów w czasie, w którym dobry szachista skupia się tylko nad kilkoma. Skąd szachista wie, że są one optymalne? Jest to kwestia równowagi pomiędzy ilością zgromadzonej wiedzy a szybkością przesuwania. Różne systemy inteligentne mają różne ograniczenia, zależnie od swojej konstrukcji. Komputery potrafią dokonywać bardzo szybkich przeszukiwań, mózgi ludzkie są pod tym względem znacznie wolniejsze. Z drugiej strony zgromadzenie obszernej wiedzy dotyczącej jakiejkolwiek dziedziny w programie komputerowym wymaga programu, który się uczy na przykładach i ma za sobą analizę bardzo wielu przykładów, lub systemu, któremu tą wiedzę podano w postaci reguł. Pierwszy sposób, tworzenie systemów uczących się, to zagadnienie nabierające w badaniach nad sztuczną inteligencją coraz większego znaczenia, ale stosunkowo nowe, rozwijane intensywnie dopiero od końca lat 80-tych, między innymi w oparciu o modele neuronowe. Drugie podejście wiąże się z systemami symbolicznymi i wymaga zgromadzenia odpowiedniej wiedzy na podstawie analizy sposobu rozwiązywania problemów przez ekspertów. Gromadzenie takich baz wiedzy jest trudnym zadaniem i stąd istniejące bazy wiedzy dalekie są od tego bogactwa wiedzy i tego wyrafinowania reprezentacji tej wiedzy ułatwiającej jej używanie, które jest udziałem człowieka. Pamięć, którą dysponujemy, pozwala nam w równoległy sposób uaktywniać tysiące reprezentacji złożonych sytuacji jednocześnie, oceniając, które z nich są najbardziej w danej sytuacji przydatne. Tylko takie reprezentacje będą następnie dalej rozważane. Konstrukcja obecnie stosowanych komputerów bardzo utrudnia podobne postępowanie i dopiero nowa generacja masowo równoległych komputerów i sieci neuronowych stanie się pod tym względem dużo bliższa działaniu mózgu.

Z punktu widzenia kognitywizmu technologia neurobiologiczna, na której oparte jest działanie mózgu, pozwala na zrealizowanie jednego typu umysłu. Technologia półprzewodnikowa pozwala na realizację umysłu innego typu, jednakże obydwa rodzaje umysłów zasługują na to miano, gdyż stanowią realizację pewnego programu i należy je rozpatrywać na poziomie wiedzy jako systemy intencjonalne. Ograniczenia systemów zrealizowanych w oparciu o różne architektury sprzętowe są z natury rzeczy odmienne, ale ich funkcja, ich sposób działania, są podobne. Stany psychologiczne człowieka tak się mają do stanów jego mózgu jak stany obliczeniowe (w sensie wykonywania abstrakcyjnego algorytmu) komputera do jego stanów fizycznych. Właśnie to twierdzenie funkcjonalizmu, określane w literaturze filozoficznej mianem **psychofunkcjonalizmu**, jest podstawą nadziei kognitywistów na stworzenie sztucznego umysłu, jest też najbardziej kontrowersyjne i najsilniej krytykowane.

1.3 Argumenty Turinga

Allan Turing w słynnym artykule „Computing Machinery and Intelligence” z 1950 roku rozważył to zagadnienie dokładnie i przeformułował problem uznając pytanie „czy maszyny mogą myśleć?” za zbyt mało precyzyjne. Zamiast tego wprowadził pewien test, określany jako test Turinga, który ma za zadanie określić, czy mamy do czynienia z inteligentną istotą czy nie. Pierwotny test polegał na próbie odgadnięcia przez człowieka, czy porozumiewając się przy pomocy terminala ma do czynienia z kobietą, czy mężczyzną. W teście biorą udział trzy osoby: A ma za zadanie zmylić pytającego tak, by nie zgadł jego lub jej płci, B ma za zadanie mu pomoc. A próbuje więc oszukiwać jak może a B przekonać go o tym, że sam/sama mówi prawdę w odróżnieniu od A. Turing proponuje wymienić A na maszynę i zobaczyć, czy pytający będą się równie często mylić w tym teście w porównaniu z sytuacją, w której A jest człowiekiem. W tej formie test jest nieco zagmatwany i obecnie za test Turinga uważa się najczęściej samą próbę udawania przez program komputerowy, że w czasie konwersacji ma się do czynienia z człowiekiem. W tej formie organizowane są od 1991 roku zawody o nagrodę Loebnera dla twórców tego systemu, który będzie w stanie oszukać komisję oceniającą, że jest człowiekiem. Temat dyskusji nie jest ograniczony, z tego powodu programy typu systemów eksperckich, wyspecjalizowane w odpowiadaniu na pytania w jednej, wąskiej dziedzinie, nie mają szans. Co roku przyznawana jest nagroda 2000 $ dla najciekawszego programu a pełna nagroda 100.000 $ przyznana zostanie dopiero po przejściu testu przez jakiś program. Na razie programy stające do tego testu oparte są na dość wąskiej bazie wiedzy i prowadzą dyskusję tylko na ściśle określone tematy.

Microsoft pracuje nad projektem o nazwie „Persona”, który jest rodzajem systemu dialogu z użytkownikiem. Graficznie przedstawiana symulacja z którą można będzie rozmawiać w języku naturalnym i która wykonywać będzie różne czynności w swoim symulowanym świecie ma stwarzać złudzenie, że mamy do czynienia ze świadomą istotą. Nawet jeśli maszyna przejdzie test Turinga, czy będziemy mogli powiedzieć, że ma ona umysł? Turing sam był przekonany, że do końca tego wieku będzie można mówić o myślących maszynach bez popadania w sprzeczności. Jego artykuł dotyczy przede wszystkim myślenia, nie wspomina o problemie uznania sztucznego systemu myślącego za umysł. Zdając sobie sprawę z tego, że opinia o możliwości stworzenia myślących maszyn jest bardzo kontrowersyjna Turing omówił szereg możliwych zarzutów.

1. Zarzut teologiczny: myślenie jest funkcją nieśmiertelnej duszy a tej maszyna mieć nie może.

Jest rzeczą ciekawą, że na swojej liście ten kontrargument Turinga umieścił na pierwszym miejscu chociaż, jak napisał w swoim artykule, nie jest zdolny do uznania tego argumentu za ważny. Najwidoczniej w 1950 roku był to w dalszym ciągu najbardziej powszechny kontrargument wśród ogółu ludności. Argumenty tego rodzaju wysuwano w historii przy każdej okazji dowodząc płaskości Ziemi, błędów Columba czy Kopernika. Odpowiedź w terminach teologicznych wygląda tu następująco: kategoryczne stwierdzanie, że maszyna nie może mieć duszy jest ograniczaniem Boskiej wolności. Podobnej odpowiedzi udzielił mi w czasie dyskusji biskup Życiński: „Nie jestem Bogiem by powiedzieć, czy maszyna może mieć duszę czy nie”.

2. Konsekwencje powstania takich maszyn byłyby zbyt straszne. Lepiej schować głowę w piasek.

Chociaż nie wygląda to na poważny argument i rzadko jest wyrażany w tej formie to przekonanie o unikalności naszego umysłu w znacznej mierze podparte jest strachem przed konsekwencjami stworzenia sztucznej inteligencji. Lepiej o tym nie myśleć i mieć nadzieję, że nic takiego nie da się zrobić. Wiele książek napisano w oparciu o takie motywacje. Na możliwości zastosowań komputerów nakładano bardzo silne bariery. Szachiści wyśmiewali programy komputerowe do momentu, gdy arcymistrzowie zaczęli z nimi przegrywać. Penrose w swojej popularnej książce „Cień umysłu” napisał, że do gry w szachy można jeszcze napisać dobry program ale do gry w **go** już nie, gdyż ma to jakoby przekraczać możliwości komputerów. Jego pierwsza książka na temat sztucznej inteligencji, „Nowy umysł cesarza”, zawdzięcza w znacznej mierze popularność naszej chęci wiary w to, że umysł jest czymś niefizycznym, niemożliwym do symulacji przy pomocy komputerów.

3. Zarzut matematyczny.

W matematyce znanych jest kilka rezultatów nakładających pewne ograniczenia na możliwości dyskretnej matematyki. Najbardziej znanym z tych rezultatów jest twierdzenie Gödla głoszące, że w systemie logicznym opartym na niesprzecznym układzie aksjomatów zawsze można znaleźć takie stwierdzenia, których prawdziwości nie da się ani udowodnić ani jej zaprzeczyć. Podobne rezultaty, odnoszące się do ograniczeń możliwości teoretycznego obliczania podane zostały przez Churcha i samego Turinga. Ma to jakoby stanowić wielkie ograniczenie możliwości sztucznego myślenia. Argument ten podnoszony jest mniej więcej raz na dziesięć lat i z powodu naszych chęci chowania głowy w piasek ciągle ma licznych zwolenników. Wspomniane już książki Penrose'a, a szczególnie “Cień umysłu” (1994), w bardzo szczegółowy sposób na 250 stronach omawia ten argument i możliwe zarzuty. Penrose dochodzi do mocnego wniosku, że stworzenie sztucznego umysłu nie jest możliwe gdyż człowiek może zawsze odpowiedzieć na pytania Gödlowskie (dotyczą one formalnej specyfikacji systemu symbolicznego, który symuluje umysł) a komputer, ograniczony przez zestaw aksjomatów na którym opiera się jego logika, tego nie potrafi. Człowiek posługuje się bowiem „świadomością”, czymś czego maszyna mieć nie może.

Odpowiedź Turinga w tym przypadku była krótka: człowiek też się myli i nie może rozstrzygnąć wielu pytań, ograniczenia naszego umysłu są bardzo silne a poczucie, że możemy odpowiedzieć na wszystkie pytania jest złudne. Rezultaty matematyczne pokazują jedynie ograniczenia konkretnej maszyny ale nie ograniczenia nieskończonego zbioru wszystkich możliwych maszyn. Dany człowiek może odpowiedzieć na pytania, z którymi nie poradzi sobie konkretna maszyna ale dana maszyna odpowiedzieć może na wiele pytań, z którymi nie poradzi sobie dany człowiek. Jeśli więc porównywać możliwości komputerów i ludzi to należy to robić dla nieskończonego zbioru jednych i drugich. Takie porównania nie są jednak tak jednoznaczne. Próba odpowiedzi przez człowieka na pytania typu Gödlowskiego nie jest możliwa choćby z tego względu, że formalna specyfikacja maszynerii naszego mózgu nie wydaje się być możliwa, a już na pewno nie prowadzi do pytania, które człowiek mógłby ogarnąć umysłem choćby ze względu na długość takiej specyfikacji.

Argument matematyczny pokazuje, ze nie można stworzyć umysłu wszechwiedzącego, takiego, który ma wiedzę niesprzeczną, doskonałą i potrafi z niej wyciągnąć wszystkie możliwe wnioski. Człowiek nie jest jednak taką istotą doskonałą. Argument matematyczny prowadzi Penrose'a na manowce, zamiast rozpatrywać możliwości modelowania konkretnych zjawisk poznawczych argumentuje on, że musimy poszukiwać zupełnie nowej fizyki procesów nieobliczalnych, czyli takich, które nie dają się algorytmizować. Nie chodzi tu o zagadnienia trudne (określane jako NP-trudne z matematycznego punktu widzenia), takie jak zagadnienia optymalizacji dla których nie ma efektywnych algorytmów, gdyż tego rodzaju zagadnienia daje się rozwiązać w sposób przybliżony. Komputery od pewnego czasu radzą sobie znacznie lepiej od nas z takimi zagadnieniami dzięki rezygnacji z metod deterministycznych szukania absolutnie najlepszych rozwiązań a zadawalaniu się rozwiązaniami bardzo dobrymi. Istnieje jednak cały szereg zagadnień, dla których udowodniono, że nie ma żadnego algorytmu

prowadzącego do ich rozwiązania, np. kwestie szukania rozwiązań układów równań Diofantycznych (posiadających całkowite współczynniki i rozwiązania będące liczbami całkowitymi). Penrose sądzi, że nasza zdolność do rozwiązywania takich zagadnień wskazuje na rolę procesów, które nie mają natury obliczeniowej w ludzkim myśleniu. Nie jest to jednak zadanie łatwe i można sadzić, ze rozwiązujemy je metodą prób i błędów posługując się wiedza o podobnych przypadkach, a więc wiedzą heurystyczną. Jest tak w przypadku obliczeń całek symbolicznych przez programy służące do algebry symbolicznej i nie ma powodu by programy do rozwiązywania równań Diofantycznych radziły sobie z tym gorzej niż ludzie. Jest to zagadnienie weryfikowalne w oparciu o metody statystyczne i obserwacje psychologiczne a nie spekulacje matematyczne.

Nie ma wątpliwości, że umysł nie działa w oparciu o ustalony algorytm, uczymy się ciągle i to w sposób nieprzewidywalny, nic więc dziwnego, że wśród miliardów umysłów pojawi się co roku kilka nowych idei. Bardzo wątpliwym jest również stwierdzenie, że naprawdę posługujemy się wiedzą niesprzeczną.

4. Świadomość, emocje, uczucia to rzeczy niedostępne maszynom.

Mechanizmu nie mogą czuć przyjemności ze swoich sukcesów, rozumieć poezji czy wpadać w depresje. Argument ten Turing zbył twierdząc, że odmowa uznania za myślącą maszyny, która przejdzie jego test prowadzi do solipsyzmu, gdyż nie mamy innej podstawy ponad to, co obserwujemy, by uznać istnienie świadomości innych osób. Powrócimy do tego argumentu bardziej szczegółowo w dalszej części tego rozdziału.

5. Argumenty dotyczące różnych niemożliwości.

Jest to cała grupa argumentów wynikających z przekonania, że maszyny nie są zdolne do samodzielnej inicjatywy, humoru, zakochania się, rozkoszowania lodami i tysiąca innych rzeczy. Takie odczucia wynikają z generalizacji naszych doświadczeń z obecnie istniejącymi maszynami i wydają się nie mieć głębszego uzasadnienia. Nie ma wątpliwości, że komputery mogą obecnie robić rzeczy, które wprawiłyby w zdumienie ludzi sprzed 50 lat. Nie mamy trudności, by uwierzyć, że komputery potrafią pamiętać więcej niż my i szybciej odszukać informacje bo od lat słyszymy i widzimy wzrastające możliwości komputerów w tym zakresie. Potrafimy już sobie wyobrazić maszyny, które będą nam udzielać odpowiedzi na poziomie encyklopedycznym na pytania zadawane normalnym głosem na prawie dowolny temat. Nasze przekonania nie są dobrymi argumentami.

6. Maszyna sam nie może nic stworzyć.

Turing nazywa argumentem Lady Lovelace' gdyż napisała ona o projekcie maszyny analitycznej Babbage'a „Nie ma ona pretensji by cokolwiek zapoczątkować samemu a robi tylko to, co się jej każe". Każdy, kto pracował nad złożonym oprogramowaniem wie, że nie robi ono tylko tego, czego się od niego spodziewamy i ciągle nas zaskakuje. Nie jest prawdą, że maszyny nie mogą wymyśleć nic nowego, gdyż nawet całkiem prymitywne systemy do dowodzenia twierdzeń z lat 60-tych podały kilka interesujących dowodów, o których nikt z ich twórców nie myślał. Programy ekspertowe odkryły wiele reguł i zależności przydatnych w licznych dziedzinach, a obecnie stosowane programy do przetrawiania danych (data mining) są w stanie na podstawie wyników pomiarów dokonać takich odkryć jak prawa Keplera czy struktura kwarkowa silnie oddziaływujących cząstek. Niezależnym problemem jest odpowiedź na pytanie: na ile twórczy jest umysł człowieka? Jest wielu ludzi, którzy nie stworzyli niczego naprawdę nowego i nikt nie twierdzi, że ludzie ci pozbawieni są umysłów. Wielkie odkrycia wynikają często z przypadkowego błądzenia. Nawet Einstein powiedział kiedyś, że człowiek nie widzi dopóki się o odkrycie nie potknie.

7. Układ nerwowy nie działa w sposób dyskretny

Von Neuman w swojej książce „Maszyna matematyczna i mózg ludzki" rozważał różnice pomiędzy urządzeniami cyfrowymi a układem nerwowym, zwracając uwagę na cechy statystyczne i analogowe impulsów. Słynna praca McCullocha i Pittsa przedstawiająca układ nerwowy w postaci sieci logicznej działającej w oparciu o algebrę Boola wywarła bardzo duży wpływ na badaczy przy końcu lat 40-tych i w latach 50-tych. Komputery analogowe konkurowały wówczas z cyfrowymi. Przybliżenia dyskretne bardzo dobrze dają się zastosować do opisu ciągłych zjawisk i obecnie komputery analogowe prawie już zniknęły z powierzchni ziemi. Cyfrowe przybliżanie zależności statystycznych również nie sprawia trudności. Urządzenia cyfrowe mogą z dowolną dokładnością przybliżyć wszystkie procesy przetwarzania informacji w układzie nerwowym.

8. Zachowanie człowieka nie da się opisać przy pomocy reguł.

Nie jest bynajmniej rzeczą jasną, ze zachowanie ludzi nie da się opisać żadnymi regułami. Z pewnością nie widać żadnego zbioru prostych, zrozumiałych dla nas reguł, które by je wyjaśniało. Nie ma jednak powodu by przypuszczać, że udało by się nam rozszyfrować skomplikowane reguły rządzące zachowaniem się złożonej maszyny, zwłaszcza jeśli jej decyzje w niektórych przypadkach byłyby przypadkowe. Jeśli funkcja celu dla wielu możliwości ma podobne wartości system ekspertowy wybiera jedną z nich w sposób przypadkowy i znalezienie reguł na podstawie obserwacji takiego systemu nie byłoby proste.

9. Argument związany z postrzeganiem pozazmysłowym.

Omówiłem już te kwestie dość dokładnie. Turing przyznaje, że jego test wymagałby wykluczenia tego typu zjawisk ale jest przekonany, że myślenie nie może mieć wiele wspólnego z tego typu zjawiskami.

1.4 Umysł jako maszyna Turinga

Modelem uniwersalnego komputera jest maszyna Turinga, zdolna do wykonywania dowolnych obliczeń. Maszyna taka może przyjąć nieskończenie wiele stanów obliczeniowych i wykonywać nieskończenie wiele programów. Oczywiście każda konkretna maszyna w skończonym czasie, podobnie jak i każdy konkretny mózg w skończonym czasie życia, może przyjąć jedynie skończenie wiele stanów ale rozważamy tu przestrzeń wszystkich możliwości. Jest rzeczą mało istotną w jaki sposób odbywają się obliczenia, jakie stany fizyczne po drodze przyjmują różne maszyny Turinga, dopóki ich odpowiedzi na zadawane pytania są identyczne, czyli ich stany obliczeniowe są identyczne, należy te maszyny uznać za równoważne. Komputery wykonują obliczenia korzystając z mikroprocesorów na poziomie manipulacji danymi w sowich rejestrach w bardzo różny sposób, używając np. różnej liczby bitów na których wykonywane są jednocześnie operacje, a jednak spełniają identyczne funkcje obliczeniowe. Właściwym poziomem rozpatrywania tych zagadnień jest więc poziom funkcjonalny, w tym przypadku obliczeniowy. Na tym poziomie wszystkie komputery „wierzą", że 2+2=4. Podobnie dzieje się ze stanami psychologicznymi: poziom obliczeniowy, poziom operacji neurofizjologicznych, nie wydaje się wpływać na nasze wierzenia, nie dostrzegamy go w naszych umysłach. Stąd przekonanie, że umysł należy rozpatrywać jedynie na poziomie wiedzy i funkcji, że poziom realizacji neurobiologicznej nie ma na niego wpływu. Te same stany psychiczne mogą być realizowane przez różne stany sprzętowe (stany mózgu lub maszynerii obliczeniowej).

Psychofunkcjonalizm mocno krytykuje teorię twierdzącą, że stany umysłowe są bezpośrednio związane ze stanami neurofizjologicznymi mózgu. Teorię taką nazywa się **teorią identyczności stanów centralnych** (central state identity theory). Gdyby tak było, twierdzą funkcjonaliści, to nie tylko podważało by to możliwość posiadania umysłu przez komputery, lecz również uniemożliwiało przypisanie umysłu istotom pozaziemskim o odmiennej od nas budowie a nawet ludziom o odmiennej budowie mózgu (np. z rozległymi uszkodzeniami pewnych obszarów mózgu). Funkcjonaliści zgadzają się, że stany psychologiczne zależą od relacji pomiędzy różnymi stanami sprzętowymi, ale nie zgadzają się, że są one z nimi tożsame. Podobnie jak behawioryści wierzą, że niezależnie od fizycznej realizacji stany umysłu dają się charakteryzować przez zależności typu bodziec-reakcja. Jak pogodzić te dwa stanowiska? Należy uznać, że ważny jest typ danego stanu, a nie konkretny stan psychologiczny czy sprzętowy. Każdy typ psychologicznych stanów można związać ze stanami fizycznymi określonego typu.

W ramach kognitywizmu rozwinęły się różne tendencje filozoficzne związane z podejściem do przekonań. **Indywidualizm** identyfikuje przekonania i sądy z wewnętrznymi stanami przetwarzania informacji. Jego formą jest **metodologiczny solipsyzm** (Fodor 1980) uznający, że z punktu widzenia nauk o poznaniu istnienie świata zewnętrznego można w pewnych celach pominąć. W szczególności dotyczy to charakteryzacji i przypisywania przekonań systemom. Pogląd taki wynika z przekonania, że jedyny dostęp do świata zewnętrznego, jaki ma system przetwarzający informację, zachodzi poprzez wiedzę i przekonania, które system posiada, nie jest to więc dostęp bezpośredni. Stany informacyjne, niezależne od stanów środowiska w tym podejściu, należy zidentyfikować, nadać im znaczenie semantyczne. W jaki sposób jednak możemy to zrobić? **Naturalistyczny indywidualizm** (Pylyshyn 1980) zaleca, by obserwować działanie organizmu w środowisku na tej podstawie przypisać znacznie semantyczne poszczególnym stanom informacyjnym. Przejście od świata fizycznego do mentalnego wymaga interpretacji zjawisk naturalistycznych. Solipsyzm metodologiczny nie pozwala na taką interpretację. W przeciwieństwie do indywidualistycznych koncepcji można uznać, że również przekonania nie są tylko kwestią systemu samego w sobie, lecz jego związków ze środowiskiem. Używanie języka zakłada np.

związek z innymi istotami używającymi języka, nie można więc wypowiedzi językowych traktować jako wewnętrzną własność systemu.

Jakości wrażeń, określane w anglosaskiej literaturze technicznym słowem „**qualia**” (od łacińskiego „qualis”, jakiego rodzaju) to inne ważne zagadnienie filozofii umysłu. Wrażenia, jakie mamy gdy smakujemy czekoladę, jakie mamy gdy obserwujemy czerwień zachodzącego słońca są trudno uchwytne i łatwo sobie wyobrazić, że mechaniczny język czy kamera wideo doznając podobnych wrażeń nie ma żadnych odczuć ich jakości. Przekonania związane są z pewnymi relacjami, tu mamy do czynienia z własnościami wewnętrznymi umysłu. Jakości są nieodłączną częścią naszego życia psychicznego, wchodzą więc w zakres zainteresowań nauk o poznaniu. Kognitywiści mają dwa różne rozwiązania tego problemu: zaprzeczyć istnieniu jakości lub twierdzić, że jakości pojawią się również w sztucznych systemach. Najprostszym rozwiązaniem, pachnącym behawioryzmem, jest zaprzeczenie istnienia problemu. Ponieważ jakości nie można scharakteryzować w oparciu o funkcjonalizm, nie mają one obserwowalnych konsekwencji, więc nie istnieją. Funkcjonalnie identyczne stany mogą się różnić jakościami wrażeń, które powodują, lecz skoro są to funkcjonalnie równoważne stany to dla funkcjonalistów różnica nie istnieje. Drugim rozwiązaniem jest stwierdzenie, że skoro mamy do czynienia z funkcjonalnie równoważnymi stanami, np. opisującymi zmęczenie czy ból organizmu, to również maszyna musi je mieć. Lepszym rozwiązaniem wydaje się być stwierdzenie, że muszą istnieć jakieś niefunkcjonalne różnice pomiędzy prawdziwymi stanami umysłowymi, w których występują jakości, a fałszywymi stanami, w których takie jakości są nieobecne. Trudno jest jednak znaleźć takie różnice a jeśli one istnieją to jakości powinny wpływać na nasze przekonania i nastawienia a więc mieć znaczenie funkcjonalne. Jeśli jednak mają znaczenie funkcjonalne to powinny się również pojawić w sztucznych umysłach.

1.5 Krytyka kognitywizmu.

Chociaż kognitywizm może wyglądać atrakcyjnie oferując nowe rozwiązanie problemu ciała i umysłu - umysł nie sprowadza się do stanów ciała, gdyż ten sam algorytm wykonywać można na rożnym sprzęcie, algorytm nie jest czymś materialnym - to w istocie nie jest to pogląd daleki od dualizmu. Umysł nie potrzebuje ciała, dowolny dostatecznie złożony sprzęt wystarczy, by go wywołać. Najkrócej program badawczy kognitywizmu ujmuje stwierdzenie: myślenie to proces przetwarzania informacji a to nic innego jak manipulacja symbolami, a więc badanie procesów obliczeniowych powinno doprowadzić do zrozumienia natury umysłu. Zadaniem nauk o poznaniu powinno być opisywanie umysłu na poziomie przetwarzania informacji, czyli na poziomie symbolicznym. Tak przynajmniej przedstawiają poglądy kognitywistów ich krytycy. Czy nie jest to jednak zbytnie uproszczenie? Program, o którym mówią kognitywiści, przetwarza przecież bardzo specyficzne dane pochodzące z naszych zmysłów, z komórek składających się na nasze ciało, z molekuł służących za neurotransmitery. Dane, z jakimi ma do czynienia, muszą mieć decydujący wpływ na jego strukturę, rozwiniętą w toku ewolucji w celu ich przetwarzania. Nie jest bynajmniej rzeczą obojętną z jakiego rodzaju mózgiem mamy do czynienia. Z doświadczenia wiemy, że nawet stosunkowo drobne różnice w budowie mózgu prowadzą do dość odmiennych stanów psychicznych - widać to po pacjentach szpitali psychiatrycznych, po odmienności świata umysłu małp człekokształtnych czy świata delfinów. Maszyna Turinga, która ma realizować program przetwarzania danych zwany umysłem musi być specyficzna i żaden robot nie będzie miał umysłu tego samego typu co człowiek. Czy będzie jednak miał jakikolwiek umysł ?

1.1.5. Chiński pokój

Jak się wydaje sam system kontrolny, pozbawiony zmysłów, ciała i środowiska, w którym się rozwija nie będzie zdolny do rozwinięcia prawdziwej intencjonalności. Amerykański filozof John Searl podał klasyczny już dziś eksperyment myślowy stanowiący argument przeciwko możliwości rozwinięcia się intencjonalności w dowolnej maszynie cyfrowej, której działanie opisywane jest w pełni przez dyskretne symbole. Opis formalny, czyli czysto syntaktyczny, jakiegoś procesu wydaje się prowadzić do niezdolności tego procesu do osiągnięcia prawdziwego zrozumienia semantyki symboli, z którymi ma do czynienia. Wyobraźmy sobie program do prowadzenia dialogu w języku chińskim, program który osiągnie wysoki stopień doskonałości, być może nawet spełni test Turinga. Czy można powiedzieć, że taki program rozumie język chiński?

Searl proponuje, byśmy wyobrazili sobie, że jesteśmy zamknięci w pokoju wypełnionym koszami z napisami w języku chińskim. Chociaż nie znamy języka chińskiego dostajemy instrukcję postępowania z regułami (w zrozumiałym dla nas języku) pozwalającymi nam manipulowanie tymi symbolami. Przez okienko ktoś wsuwa nam nowe symbole zawierające pytania. Analizujemy te symbole i posługując się regułami zestawiamy z

symboli wziętych z różnych koszy odpowiedzi. Jeśli reguły, odpowiadające programowi manipulacji symbolami, są dostatecznie dobre, to odpowiedzi wysuwane przez nas przez okienko będą dla stojących po drugiej stronie chińczyków miały sens i będą oni skłonni założyć, że w środku musi być ktoś, kto rozumie zadawane mu pytania. My jednakże nie mamy pojęcia o znaczeniach symboli do nas docierających, w oparciu o reguły zestawiamy jedynie symbole. W pokoju nie ma więc nikogo, kto mógłby rozumieć język chiński. Jeśli jednak wykonywanie formalnego programu nie jest wystarczającą przyczyną dla nas by rozumieć, co robimy, nie może być też mowy o tym by maszyna wykonująca ten program naprawdę coś rozumiała. Rozumienie języka wymaga czegoś więcej niż operacji na symbolach przy pomocy reguł formalnych. Jednakże maszyna cyfrowa niczego poza formalnymi symbolami znać nie może. Searl wnioskuje stąd iż jest oczywistym, że sama syntaktyka nie wystarcza do semantyki.

Podano wiele odpowiedzi na ten eksperyment myślowy: rozumienie pojawi się w systemie jako całości, rozumienie pojawi się wtedy, gdy cały system będzie we wnętrzu robota wyposażonego w zmysły i działającego w świecie rzeczywistym, do rozumienia konieczne jest używanie systemów wieloprocesorowych i wiele innych. Żadna z tych odpowiedzi nie wydaje się jednak zadawalająca tak długo, jak długo mamy do czynienia z maszyną cyfrową, gdyż nie ma sposobu by przyporządkować symbolom przez nią używanym znaczenia. Sam program i podawanie odpowiednich symboli w odpowiedzi na pytania, nawet jeśli działa w inteligentny sposób na najwspanialszym komputerze nie gwarantuje jeszcze rozumienia i myślenia. Komputer może symulować myślenie ale nie naprawdę myśleć. Wniosek z tego eksperymentu myślowego prowadzi więc do odrzucenia testu Turinga oraz całej idei kognitywizmu, która może doprowadzić do symulacji inteligencji ale nie do sztucznego umysłu. Oddzielając umysł od procesów biologicznych, realnych stanów materii, zastępując go natomiast programem, funkcjonaliści nadają mu status bytu odmiennego rodzaju, oderwanego od świata biologicznego. Przyjęcie radykalnej wersji sztucznej inteligencji jest więc formą dualizmu. Searl odrzuca taki pogląd proponując w zamian następujące rozumowanie. Prawdziwe są następujące cztery przesłanki:

1.Mózgi są przyczyną umysłów
2.Syntaktyka nie wystarcza do semantyki
3.Program komputerowy całkowicie określa syntaktyka
4.Umysły zawierają treści semantyczne (psychiczne)

Stąd wypływają następujące wnioski:

W1. Programy nie wystarczają do powstania umysłu.
W2.Czynności mózgu ograniczone do realizowania programów komputerowych nie wystarczają do powstania umysłu.
W3.Przyczyna powstania umysłu musi mieć porównywalną moc oddziaływania przyczynowego z możliwościami mózgu.
W4.Wyposażenie robota czy innego sztucznego systemu w program komputerowy nie umożliwi mu posiadania stanów umysłowych porównywalnych z ludzkimi. Stany umysłowe są zjawiskiem biologicznym.

Dlaczego kognitywiści uznali, że można wprowadzić poziom symboliczny abstrahując od poziomu procesów neurofizjologicznych? To właśnie ten krok oddziela nas od neurobiologii, realnych stanów mózgu z którymi mamy do czynienia u ludzi, a pozwala zastąpić je symbolami i przetwarzaniem informacji. Według Searle'a głębsza przyczyna wiary w powodzenie programu kognitywistycznego wypływa z analogii pomiędzy obserwacją zachowania się ludzi postępujących w oparciu o reguły i komputerów, również działających w oparciu o reguły. Skoro i ludzie i komputery przestrzegają pewnych reguł to może działają według identycznych zasad. Nie jest to jednak dobra analogia. W przypadku ludzi postępowanie zgodnie z jakimiś regułami wynika ze zrozumienia sensu reguł a nie z konieczności wykonywania programu. Komputery kierują się regułami jedynie w metaforycznym sensie tych słów, podobnie jak kieruje się regułą Słońce wstające na wschodzie i znikające na zachodzie, komputery wykonują jedynie krok po kroku swoje programy.

Jak więc widać Searl daleki jest od uznania, tak jak czyni to Newell, że na poziomie wiedzy uprawnione jest używanie w stosunku do systemów opartych na wiedzy języka metaforycznego. Podobnie procesy przetwarzania informacji w komputerach należy rozumieć w innym sensie niż przetwarzanie informacji u człowieka, np. dodając dwie liczby maszyna nie wie, co oznacza 2+3, i dzięki swojej niewiedzy może sprawnie i szybko dokonać wyspecjalizowanych działań (ciekawe, że „idiot savants", czyli osoby o nadzwyczajnych zdolnościach w jakimś wąskim zakresie, też wydają się mieć zawężone rozumienie pojęć, np. liczb na których wykonują bardzo szybko operacje). Symulacja pewnych aspektów działania umysłu powoduje, że traktujemy programy tak, jakby ich przetwarzanie było równoważne umysłowemu, gdyż przyczyny i skutki są takie same. Różne

procesy fizyczne opisać można tak, jakby przetwarzały one informacje, np. promień światła załamuje się w wodzie tak, jakby chciał zminimalizować czas potrzebny na dotarcie do dna. Mózg zachowuje się więc tak, jakby przetwarzał informację i dla przybliżenie opisu jego działania takie porównanie może być całkiem użyteczne.

Searle uważa, że szanse na powodzenie programu kognitywistycznego są zerowe i zakłada, że pomiędzy stanami mózgu a stanami psychicznymi nie ma żadnego pośredniego poziomu. W dosłownym sensie jest to prawda ale raczej nieciekawa, gdyż nie pozwala nam ona na rozwinięcie teorii naukowych, które zawsze opierają się na upraszczaniu i przybliżaniu. Podobnie można by powiedzieć, że pomiędzy zachowaniem się komórki a elementarnymi procesami na poziomie jąder i elektronów nie ma żadnego pośredniego poziomu: materia zachowuje się tak, jak gdyby były atomy, makrocząsteczki, reakcje chemiczne, ale naprawdę to tylko oddziaływania elektromagnetyczne pomiędzy elektronami i protonami. W ten sposób rezygnujemy z całej chemii. Płodny punkt widzenia wymaga wprowadzenia wielu poziomów pośrednich jeśli tylko uda się znaleźć na wyższym poziomie stabilne formy (takie jak atomy i cząsteczki chemiczne złożone z jąder i elektronów) i reguły oddziaływania pomiędzy nimi. Opis umysłu jest kwestią odpowiednich aproksymacji, stworzenie umysłu wymaga użycia prawdziwej materii.

Wydaje się, że argument chińskiego pokoju podobny jest do argumentów Gilberta Ryle głoszących konieczność inteligentnych procesów do wyjaśnienia inteligencji i w efekcie prowadzących do „ducha w maszynie”. W jaki sposób moglibyśmy stwierdzić, że w mózgu jakiegoś człowieka zachodzą procesy umysłowe? Na pewno nie przez umieszczenie tam demona, rozpoznającego sygnały czy impulsy nerwowe. Demon działający lokalnie nie może objąć całości, nie może odczuć, że mózg jako całość ma pewne stany. Możemy krytykować pomysł Turinga i uznać, że jest on niewystarczający jako test na intencjonalność, ale czy chiński pokój Searla jest takim testem? Dobry test powinien pozwolić na jednoznaczne stwierdzenie, że w przypadku człowieka mamy do czynienia z umysłem a w przypadku komputera tylko z przetwarzaniem informacji, test Searla tego jednak nie potrafi. Jedyny sposób na stwierdzenie, czy mamy do czynienia z umysłem to próba synchronizacji naszego mózgu z systemem zewnętrznym, próba zestrojenia wibracji elektrycznych dwóch struktur. Test taki nie jest praktycznie możliwy, a gdyby był, nadawał by się tylko do wykrycia umysłu o strukturze bardzo zbliżonej do naszej. Nie dysponując bowiem odpowiednimi fragmentami kory mózgu nie będziemy w stanie odczuć i zrozumieć sygnałów z sonaru delfina czy echosondy nietoperza.

1.2.5. Nieobliczalność

W odróżnieniu od poważnej krytyki filozoficznej przedstawionej powyżej zarzuty wysunięte przeciwko kognitywizmowi przez Rogera Penrose'a w jego popularnej książce „Shadows of the mind” są nieporozumieniem. Wspominam o niej tylko dlatego, że na pewno narobi ona dużo szumu i wywoła wiele zachwytów - w prasie polskiej już w 1995 roku ukazało się kilka recenzji traktujących ją jako książkę ważną, wymagającą głębokich dyskusji. Książka ta, podobnie jak i wcześniejsza książka Penrose'a „Nowy umysł cesarza”, zawdzięcza swoją popularność bardziej chęci chowania głowy w piasek, jak to pisał Turing, niż swojej merytorycznej wartości. Penrose jest znanym matematykiem pracującym również nad zagadnieniami fizyki matematycznej. Wyciągnięty przez niego argument oparty na twierdzeniu Gödla i Turinga nie jest nowy i roztrząsanie go na 250 stronach pokazuje, że nie jest bynajmniej prosty. Penrose pisze przede wszystkim o świadomości. Na początku książki stwierdzając, że „świadomość bez wątpienia jest czymś” nie usiłuje wcale zdefiniować, o jakie konkretnie zachowania czy zjawiska poznawcze mu chodzi. Wyróżnił on cztery podejścia do świadomości:

A) Myślenie jest po prostu obliczaniem a świadomość wynikiem tych obliczeń.

B) Świadomość jest cechą fizycznych właściwości mózgu. Można je symulować rachunkowo ale samo obliczanie nie powoduje świadomości.

C) Fizyczne działania mózgu wywołujące świadomość nie mogą być symulowane rachunkowo.

D) Świadomości nie można wyjaśnić metodami naukowymi.

Stanowisko A jest rezultatem zgody na test Turinga nie tylko do oceny inteligencji ale i do oceny, czy mamy do czynienia z umysłem świadomym. Właśnie to stanowisko Penrose krytykuje ostro ale bez zrozumienia, przypisując jego zwolennikom zupełnie dziwaczne intencje. Roztacza przed czytelnikami wizje algorytmu

umysłu zapisanego w książce i zadaje pytanie, w którym momencie ma się pojawić umysł - czy od samego spoczywania w książce czy może przy przewracaniu jej kartek? Niestety, nikle pojęcie autora o istocie badań w zakresie sztucznej inteligencji jest wyraźnie widoczne i w znacznej mierze książka ta walczy z wiatrakami poszukując jakiegoś umysłu abstrakcyjnego, nie związanego z konkretnymi zjawiskami badanym przez psychologów poznawczych. Penrose ledwo zauważa, że człowiek ma mózg, nie podając nawet podstawowych faktów dotyczących związku mózgu z wyższymi czynnościami poznawczymi. Jego interpretacje doświadczeń nad świadomą percepcją są całkiem błędne. Twierdzi, że nie da się zrobić dobrego programu do gry w go - podobnie jak jeszcze dziesięć lat temu wyśmiewano się z programów do gry w szachy.

Penrose pomija zupełnie stanowisko B (podobnie jak i D), skupia się natomiast na dowodzeniu, że z twierdzenia Gödla i Turinga wynika, że maszyny nie mogą rozwiązać zagadnień niealgorytmicznych, gdyż ich rozumowanie oparte jest na regułach, ludzie natomiast mogą takie problemy rozwiązywać, gdyż posługują się „świadomym myśleniem". Nie ma na to co prawda dowodów poza stwierdzeniem, ze w pewnych zagadnieniach wystarczy popatrzeć i widać rozwiązanie geometryczne zagadnienia. Analiza obrazu nie jest jeszcze na tak zaawansowanym etapie by można było korzystając z obrazu kamery wywołać odpowiednie ślady pamięci i skojarzenia w komputerze, nie ma jednak powodu, by miało to być „w zasadzie" niemożliwe. Penrose bardzo często odwołuje się do tego „w zasadzie", w odróżnieniu od „w praktyce", gdyż łatwo widać, że w praktyce nie będzie trudno obejść piętrzonych przez niego trudności, jest to tylko kwestia złożoności systemów komputerowych. Połowa książki poświecona jest poszukiwaniu ezoterycznej, nowej fizyki potrzebnej jakoby do zrozumienia umysłu. Penrose puszcza tu całkowicie wodze fantazji twierdząc, że to grawitacja kwantowa, teoria która być może ktoś sformułuje w bliżej nieokreślonej przyszłości, będzie pomocna w zrozumieniu natury umysłu. Wydaje się to skrajnie nieprawdopodobne gdyż, jak już wspominałem, fizyka kwantowa jest zbyt trudna by zastosować ją do opisu pojedynczych komórek lub nawet cząsteczek białek, nie wspominając o tak złożonej strukturze jak mózg. Książka Penrose'a sprowadza zagadnienia rozumienia działania umysłu na zupełne manowce sugerując, że konieczne jest zrozumienie roli mikrotubul, struktur cytoskeletalych, znajdujących się we wszystkich komórkach, nie tylko w komórkach nerwowych. Można odnieść wrażenie, że miliardy neuronów w mózgu znajdują się tam przypadkowo i nie mają wiele do roboty.

1.3.5. Inne zarzuty

Można wyobrazić sobie wiele eksperymentów myślowych usiłując zrozumieć naturę umysłu. Przypuśćmy, że będziemy po kolei zastępować poszczególne obszary mózgu przez sztuczne, krzemowe elementy dostarczające do obszarów nieuszkodzonych sygnałów nieodróżnialnych od prawdziwych. Już teraz rozważa się np. wszczepianie sztucznych, krzemowych siatkówek ludziom, którzy utracili wzrok. Postępując tak krok po kroku zamienimy cały mózg na urządzenie krzemowe, działające w sposób cyfrowy w oparciu o wysyłanie impulsów. Zgodnie z Johnem Searle w pewnym momencie tego procesu umysł zniknie, przestanie przynajmniej naprawdę odczuwać, zniknie jego rozumienie. W którym momencie ma się to stać? Czy zastąpienie kory wzrokowej przez krzemowy odpowiednik spowoduje brak zrozumienia wrażeń wzrokowych?

1.6 Podsumowanie

Wydaje się, ze znaczną część dyskusji na temat kognitywizmu i ograniczeń sztucznej inteligencji nie rozróżnia pomiędzy inteligencją a bardziej subtelnymi problemami dotyczącymi natury umysłu. Zarzuca się np. Turingowi, że zakłada on iż system nerwowy jest maszyna cyfrową. Artykuł Turinga miał na celu przedstawienie testu pozwalającego na określenie, czy maszyny mogą myśleć, nie przesądzał jednak bynajmniej, czy to myślenie będzie związane z intencjonalnością czy też tylko będzie symulacją ludzkiego myślenia. Nie widać powodów, dla którego inteligencja i myślenie miałyby być równoznaczne z posiadaniem stanów umysłu pozwalającym na odczuwanie wrażeń czy intencjonalność. Warto popatrzeć na to zagadnienie z punktu widzenia możliwości aproksymacji. Z roku na rok programy do analizy mowy człowieka wypowiadanej w sposób ciągły stają się coraz doskonalsze, programy do odczytywania pisma ręcznego również. Komputery sortują pocztę rozpoznając kody pocztowe znacznie szybciej niż ludzie. Podobnie jak z grą w szachy tego typu zdolności człowieka prędzej czy później przestaną być czymś nadzwyczajnym i komputery zrobią to lepiej. Już w tej chwili istnieją programy, których zdolność do rozumienia subtelności konstrukcji językowych przekracza możliwości wielu ludzi. Trudno jest wskazać konkretny przykład, w którym ciągły, chociaż nie zawsze szybki, postęp nie doprowadzi w końcu do przekroczenia możliwości człowieka. Do niektórych zadań komputery dopiero dorastają ze względu na stopień złożoności, konieczny do ich wykonania. Nie zapominajmy, że człowiek ma do dyspozycji 10^{14} parametrów ulegających adaptacji (połączeń synaptycznych pomiędzy

neuronami), podczas gdy nawet najlepsze obecnie komputery są tysiące, jeśli nie dziesiątki tysięcy, razy prostsze. Dopiero od niedawna uczenie się maszynowe nabrało rozpędu, pojawiły się całe nowe gałęzie wiedzy takie jak algorytmy ewolucyjne czy sieci neuronowe. Wszelka ocena ograniczeń maszyn wydaje się więc przedwczesna.

Właściwie postawione pytanie brzmi: czy aproksymacja możliwości umysłu człowieka przez coraz doskonalsze maszyny prowadzi jedynie do sztucznej inteligencji, czy też do sztucznego umysłu? Inteligentne zachowania da się symulować coraz lepiej ale czy w cyfrowym umyśle naprawdę ktoś mieszka? Maszyna, która spełni test Turinga, nie musi być wcale prawdziwym umysłem, może być jedynie jego doskonałą symulacją. Dźwięk dobiegający z płyty kompaktowej wydaje się nam doskonałym złudzeniem rzeczywistości. Gdybyśmy jednak z przyczyn technicznych mogli zamiast 100.000 bitów w ciągu sekundy odczytywać tylko 10.000 dźwięk zmieniłby się nie do poznania i z pewnością wielu krytyków stwierdziłoby, że wszelkie próby symulowania wrażeń czy głosów są z góry skazane na niepowodzenie. Czy powstanie umysłu jest tylko kwestią jakości przybliżenia?

Z drugiej strony trudno jest odmówić racji krytyce radykalnej wersji sztucznej inteligencji Johna Searla. Stany mózgu są fizycznymi stanami materii a nie tylko przepływem informacji. Swiadczą o tym przypadki uszkodzeń neurologicznych, np. zniszczenia części mózgu przez wylewy. Uszkodzenie fragmentów kory wzrokowej odpowiedzialnej za przetwarzanie koloru powoduje zanik zdolności do wyobrażenie sobie koloru, przezywania wrażeń z tym związanych, nawet w snach kolor przestaje się pojawiać. Oliver Sacks opisuje przypadek człowieka, który w czasie snu doznał rozległego wylewu pozbawiającego go całkowicie ośrodków wzrokowych. Wybikiem była niezdolność do wyobrażenia sobie wrażeń wzrokowych. Wrażenia, zdolność czucia, wymagaja więcp osiadania odpowiednich strukut mózgu. Moje wyobrażenia mają charakter jakościowy dzięki temu, że pobudzają do określonego stanu fizycznego - widać to w sygnałach EEG czy obrazach NMR - odpowiednie fragmenty kory mózgowej i ośrodków podkorowych. Bez takich struktur czucie, jakości wrażeń, nie istnieją.

Przypuśćmy, że chcemy odtworzyć wibracje skomplikowanej membrany. Możemy ten problem dokładnie przeanalizować przy pomocy komputera i oglądać na ekranie monitora różne możliwe wibracje do jakich membrana jest zdolna. Możemy przewidywać kolejność pojawiania się różnych stanów wibracyjnych, przepływ energii w prawdziwym układzie, będziemy mówić o stanach wibracyjnych modelu. Czy nasz program stanie się w ten sposób membraną? Niezależnie od stopnia szczegółowości naszych symulacji, nawet jeśli będziemy symulować ruch pojedynczych atomów, uzyskamy coraz dokładniejsze przewidywania, doskonały model, który może pozwolić na szybszą odpowiedź dotyczącą wszystkich aspektów zachowania się membrany niż badania doświadczalne. Symulacja nie stanie się jednak nigdy rzeczywistością. Nawet najdoskonalsza symulacja funkcji mózgu nie zamieni się w prawdziwy mózg, przyjmujący prawdziwe stany fizyczne i doznający prawdziwych wrażeń.

Literatura

Rene Descartes, *Rozprawa o metodzie*

Rene Descartes, *Medytacje o pierwszej filozofii.*

David Hume, *Badania dotyczące rozumu ludzkiego*

David Hume, *Traktat o naturze ludzkiej*

Immanuel Kant, *Krytyka czystego rozumu*

John Locke, *Rozważania dotyczące rozumu ludzkiego*

Allen Newell i Herbert Simon, *Computer science as empirical enquiry.* Communicattions of the ACM 19 (1976) 113-126

Roger Penrose, Shadows of the mind: a search for the missing science of consciousness (Oxford University Press 1994)

Platon, *Państwo*

Richard Popkin, Avrum Stroll, *Filozofia* (Zysk i Ska, Poznań 1994)

Oliver Sacks, CZłowiek, którypomylił swoją żonę z kapeluszem (Zysk i Ska 1996)